

SOCIAL NETWORK ANALYSIS AND DIGITAL LEARNING ENVIRONMENTS: A FRAMEWORK FOR RESEARCH AND PRACTICE USING THE SAPO CAMPUS PLATFORM

Luís Pedro¹, Carlos Santos¹, João Batista¹, Guilherme Cabral¹, Fátima Pais¹,
Conceição Costa²

¹ Universidade de Aveiro (PORTUGAL)

² Universidade Lusófona de Humanidades e Tecnologias (PORTUGAL)

Abstract

The main purpose of this study is to explore the research underlying the development of Social Network Analysis (SNA) features in the SAPO Campus platform, aiming to provide a set of analysis tools that can help managers, teachers and researchers to understand and react timely on the evolution of their communities. The paper presents the SAPO Campus digital platform, and the basic principles of SNA are described. To justify the rationale underlying the selection of SNA's features, a use case is described. The case is about using those features to analyse a dynamic social structure: The Department of Communication and Arts (University of Aveiro, Portugal) space in SAPO Campus. Tools such as Gephi were used to collect and format data into a graph network. The collected data refers to the period 2012-2015, with 1,001 registered users in the platform. SNA features were applied to analyze the network in terms of cohesion, degree, centrality, and group structure. The findings indicate that the analyzed network corresponds to the organization in scientific areas in DeCA, and the results found echo the perception that the authors have of the platform use. Also, results show the importance of some teachers in the promotion and dynamics of the network. These findings help to justify the use of SNA to analyze networks of this nature. However, the authors acknowledge that a more detailed interpretation of the results depends on deep knowledge of the studied network connections, suggesting that the implementation of SNA must be triangulated with complementary application of qualitative techniques. This study provides valuable insights to the development of the SNA export tool for the SAPO Campus platform under the GAMILearning project, and to a more general application of the SNA techniques as well.

Keywords: SNA, Learning Environment, Higher Education, Social Web, Education, SAPO Campus, Innovation, Technology.

1 SAPO CAMPUS: A COMMUNICATION PLATFORM FOR EDUCATIONAL AND TRAINING CONTEXTS

SAPO Campus¹ (SC) is a digital communication platform that aims to promote a set of principles related with openness, sharing, and collaboration, recognising the social dimension of knowledge construction and its expression in educational and training communities. As a technological platform, SAPO Campus implements these principles through its core features that underline:

- a dismissal of the traditional roles assigned to members of educational communities (all users have the same privileges and responsibilities in this platform) and the possibility of following, in the platform, the users with whom there is an identification in terms of learning interests and the contents they share;
- a promotion of different forms of communication and expression, providing different features that allow users to post status, comments, photos, videos and blog posts and the possibility to use digital badges in order to implement distributed and more democratic forms of assessment or to, simply, promote motivational strategies in a gamified platform;
- the possibility of creating groups that can serve as spaces for enrolling the platform users in different and/or complementary activities more related with their informal learning preferences.

This platform is being successfully used in several teaching levels (ranging from elementary schools to higher education institutions) and in several countries [1][2][3][4].

¹ http://campus.sapo.pt/?lang=en_us

Recent research on the use of this platform also indicates that social interaction in the platform is an affordance that should not be neglected in terms of its learning potential [5]. These results clearly indicate that, in order to better understand the platform usage patterns and its influence and potential in the teaching and learning process, there should be an investment on the development of social network analysis instruments that will allow a continuous monitoring of the social activity in SAPO Campus.

2 SOCIAL NETWORK ANALYSIS: RESEARCHING THE SOCIAL DIMENSION ON NETWORKED PLATFORMS

The Social Network Theory states that a network evolves as the actors develop some kind of links, either formal or informal, related to friendship or trust, influence, recognition, work, knowledge or other forms of social relation [6][7]. The social network paradigm incorporates a strong mathematical and statistical foundation in a program of cumulative research on the properties of social networks [8].

By the end of the 20th century, social network analysis had become a mode of structural analysis with extensive formal techniques at its disposal [9].

Social Network Analysis (SNA) analyses patterns of relationships between actors in a network. This methodological approach has been used in several areas enhancing the ways of examining intra-group formal and informal relationships and analysing individual attributes [10][11]. In a social group, there are relationships and ties to be considered between subjects, such as social cohesion, social preferences, social pressure or social hierarchies.

A social network consists of a finite set or sets of actors and the relations defined on them [9: 20]. In case of human social networks, actors refer to individual persons. Depending on the source of data, relations in a social network could come from a verbal or written communication and physical proximity or only virtual relations. More, taking into account that learning processes nowadays are increasingly diverse in terms of contexts (formal and informal), modality (face-to-face, blended, at distance), and rely heavily in communication and interactions between agents, often using technological tools along those processes, social network analysis will be of the utmost importance in analysing online learning communities. In a learning and research project at elementary school that involved face to face and online learning activities, SNA combined with participant observation was of great value to measure peer-to-peer and students-content interactions and their contribution to learning [12].

SNA considers three core elements – the network, the interactions and the attributes. The network is a group of people that interact together for some reason and these individuals are actors represented graphically by nodes. The interactions between them are represented by links. Each tie might be directed, undirected or bonded. Attributes are individuals' features that are able to influence relationships inside a group: age, skills, status, language, experience, among others [11][13]. The data to perform SNA can be collected through different forms - surveys, questionnaires, observation, written documents analysis or online existing data sources.

SNA has a vast number of social network concepts that are important to the analysis comprehension. The core SNA concepts relevant to this paper are described in the section 3.

3 ANALYSING THE SOCIAL STRUCTURE OF A HEI DEPARTMENT: THE DECA CASE

3.1 The Department of Communication and Arts (DeCA)

The Department of Communication and Arts (DeCA) of the University of Aveiro was officially founded in 1996 having a transdisciplinary mission of criss-crossing several scientific fields from the areas of Communication, Design, Art Studies and Music. In terms of formative options, DeCA currently offers three undergraduate programmes (Design, Music and New Communication Technologies), five Master programmes (Music, Music Education, Design, Contemporary Artistic Creation and Multimedia Communication) and four doctorate programmes (Music, Design, Information and Communication in Digital Platforms and Multimedia in Education).

The Department has 90 professors, 1226 regular students and 7 staff members.

In terms of distribution of students by scientific areas, in the 2014/2015 school year the Communication area had 492 students, the Design area 342 students, the Music area 366 and the Art Studies area 26 students.

3.2 SAPO Campus: The DeCA use case

SAPO Campus was introduced at DeCA in an early stage of the development of the platform. During the 2012/2013 school year, a total of 12 educational institutions were invited to be pilot schools with an early access to the platform. Pilot schools started to use the platform in a very basic setup, presenting only basic functionalities. Early users from different institutional levels (schools' directors, teachers and students) were an important part of the research process contributing with a stream of feedback that helped the R&D to fine tune the platform and also to define roadmap priorities based on real users' opinions.

School pilots included institutions ranging from the elementary level to a university scenario. The DeCA scenario was a relevant one because R&D team leaders are members of that institution and they also made part of the pilot phase by promoting and integrating the platform in a set of formal courses of different levels: degree, master and doctoral studies.

The results presented in this paper should be interpreted taking in consideration that during this time there was not a formal institutional support for the use of the platform. Technology was made available and it has been only promoted by cross talks between teachers, researchers and students.

3.3 Data processing methodology

3.3.1 Data Gathering

SC data related to users and content is stored in a graph database supported with Neo4j², which does not provide a way to export data directly into Gephi³, the main SNA tool used for this research. So the information about the users and the relations between them was exported to a Comma-Separated Values (CSV) text file. Each line of this file represents a relationship between two users and contains the information about the source and the target user of that relationship, the content and relationship types, the date each user joined the platform and the date that relationship was established. Two types of relationship in SC were considered. Direct relations establish connections between the source and destination of each action. Invitation relationships or following someone else's activity are examples of direct relations. Indirect relations require a subjective choice of the researcher as, for instance, assuming that a publication in a group creates a relationship with all members of that group, or considering that a comment in a publication only creates a relationship with its author, and not necessarily with the authors of the previous comments. A similar approach was adopted to describe the relationships that are established within each group.

3.3.2 Formatting data

To import the data from the SC graph database, the nodes file contained the information about all the users - both source users and target users - in the list exported from the database. The joining dates were also added for each user, so that, later, it would be possible to analyze the network growth over time. The edges file was identical to the original file exported from the database. Although the source and target columns were the only ones required to Gephi, some of the data from the database was also inserted in the edges file, for instance, the date the relationship was established and the columns about the types of content and relationship. Two more columns were added: the weight of the relationship, which was filled with the value "1"; and a column named "Type", filled with the word "directed". This column defines if the relationships are unilateral or bilateral. This two files were saved in CSV format.

3.3.3 Interpreting data on Gephi

After importing the nodes and edges files to Gephi, a graph network was generated. When the edges file is imported, all the repeated relations (for example, a user can comment another user's post several times) are merged into one and its weights are summed.

² <http://neo4j.com>

³ <https://gephi.org>

At this initial stage, a graph is generated where all the nodes' positions are random, but Gephi provides several tools to make the graph more readable. It has some layout algorithms that can spread the network. In this case, the Force Atlas algorithm was used and tweaking some of the options, the network graph got readable.

The readability of the graph was increased with the help of some other tools. It is possible to change the aspect of the nodes themselves according to the data of the network. The nodes were scaled according to each one's degree and, after calculating the Modularity, each community was colored differently.

4 RESULTS

4.1 Cohesion

After gathering and processing the data related to DeCA, the graph structure was created. By the end of December 2015, there were 1,001 users registered in the DeCA space in SC, and these users established a total of 67,107 relationships between them. Fig. 1 represents the sociogram of the DeCA space since January 2012 until December 2015.

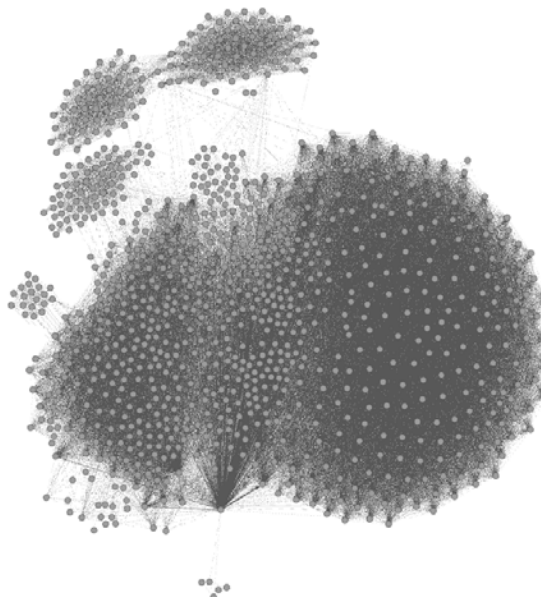


Figure 1 General Sociogram of DeCA

This network has a diameter of 7 and the average path length is 2.53. It has a density value of 0.06%, which is considered to be a low density, sparse network, as seen in the above sociogram. The eccentricity analysis of each node is represented in Table 1.

Table 1 Eccentricity

Eccentricity	Number of Nodes	Percentage of Nodes
0	94	9.39%
1	3	0.30%
2	0	0.00%
3	0	0.00%
4	3	0.30%
5	427	42.66%
6	450	44.96%
7	24	2.40%

A node that has a eccentricity of 0 is a node that has never started a relationship but was targeted by another relationship initiated by another node. On the other hand, the nodes with the eccentricity of 7 are only sources of relationships and were never targeted by any other relationships. Considering the data in the table above, the DeCA network presents a high percentage of nodes in the upper eccentricity values that indicate that a rather low level of cohesion.

4.2 Degree

To analyze the network considering the centrality of each actor, the metrics used were degree, betweenness, closeness and eigenvector, which allow to determine the position of each actor on the network.

The network has an average weighted degree of 362.77, and the distribution of different degrees are represented in Fig. 2.

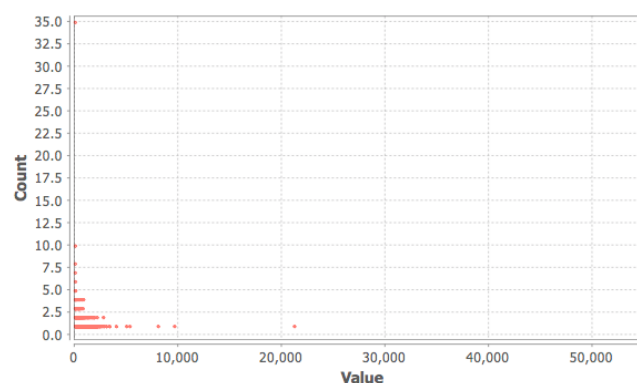


Figure 2 Degree distribution of relationships

The chart shows a concentration of nodes with a degree below 5,000, with several nodes with a degree value near to 0. Fig. 3 illustrates a detailed view of the nodes' in and out-degree.

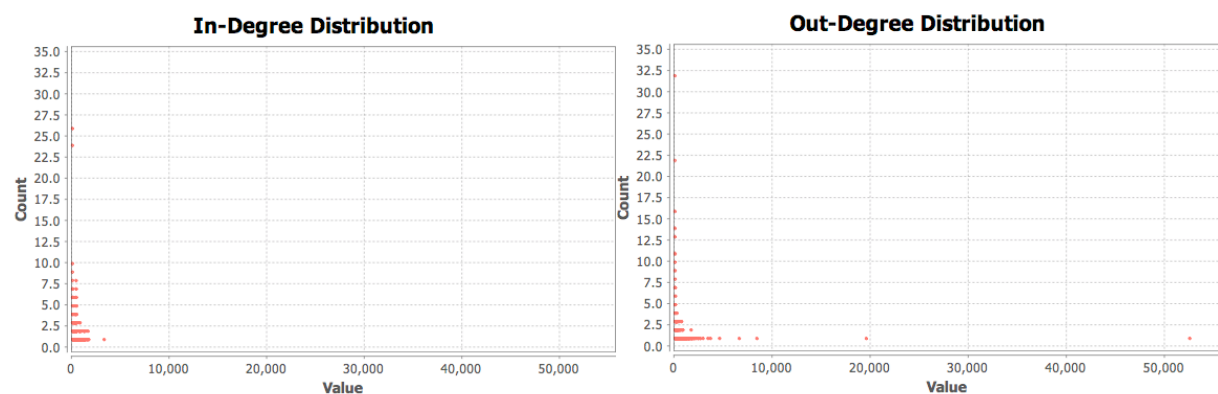


Figure 3 Weighted in-degree (left) and out-degree (right) distributions of relationships

Despite the higher concentration on the lower values, the chart shows a few nodes that detach from the majority, having higher out-degree values. Table 2 shows the degree distribution with a more in-depth view and also dispersion measures.

The analysis of the table above indicates that a large majority of users, almost 80%, has a degree value of 500 or less. These results point to the fact that there are some users that have a degree a lot higher than the majority, as already shown previously. Two measures of dispersion were calculated, standard deviation and variation coefficient. Results show that there is a big dispersion on the degree distribution in the network, especially in the out-degree. This means that there's a group of specific users that established several strong relationships, while the majority of users established weaker relationships.

To better understand the balance between teachers and students' degrees in the network a list of the 20 users with higher values for the degree measures was elaborated (Table 3). The degree list shows that the four top users are all teachers and that in the remaining list there is only one more teacher at the 15th position. The top two positions of the in-degree measure are also for teachers but, with the exception of the top user (427), the gap until the end of the list is very short (270 to 233) and there is only one more teacher at position 12. On the other hand, the out-degree measure shows that five of the top six positions belong to teachers, making a significant difference between the relevance of teachers in the in-degree and out-degree measures.

Table 2 Frequency and measures of dispersion of weighted in and out-degree distribution

	Interval	In-Degree		Out-Degree	
Frequency	[0-500[788	78.7%	795	79.4%
	[500-1000[147	14.7%	138	13.8%
	[1000-1500[58	5.8%	40	4.0%
	[1500-3500]	8	0.8%	22	2.2%
Dispersion	Average	143		91	
	Standard Deviation	289.5		300.5	
	Variation Coefficient	2.0		3.3	

4.3 Centrality measures: closeness, betweenness and eigenvector

Closeness defines that a node is more central as it has the same distance for any other node. Fig. 4 (left) illustrates the closeness values on the DeCA network,

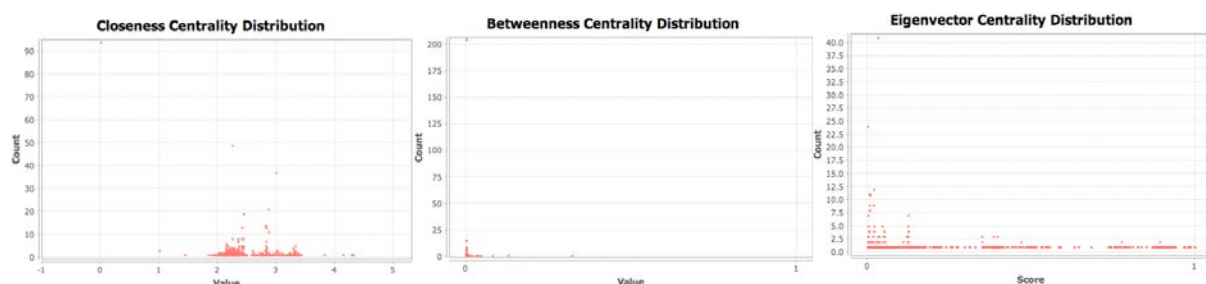


Figure 4 Closeness, betweenness and eigenvector centrality distribution

Like eccentricity, the closeness value of 0 represents the nodes that never started a relationship but were target of relationships initiated by other nodes. As Fig. 4 (left) illustrates, a large number of actors never started a relationship. This measure must be read inversely: the lower the closeness value of a node, the more central it is. So, excluding the null values, the majority of nodes has a closeness value between 2 and 3, which is reasonable, considering the diameter of the network.

Betweenness defines each node centrality as its capacity to intermediate relationships with other nodes. Fig. 4 (middle) illustrates the betweenness distribution on the network. In this case, 205 actors have a null value of betweenness, which means that 20.5% of users never intermediated a relationship. On the other hand, only two users have a betweenness centrality value of 0.08 or higher.

The eigenvector measures the influence of a node in a network. Nodes with an eigenvector centrality value closer to 1 are the most influent users. As show in Fig. 4 (right), there is a higher concentration of nodes with lower values of eigenvector centrality. Only 8.1% of actors have an eigenvector centrality value of 0.75 or higher. On the other hand, 89.6% of actors have an eigenvector centrality vector of 0.5 or lower. There are also 24 nodes with a value of 0, which means that these nodes where never target of any relationship. These results lead to the conclusion that users don't start relationships with other influent users.

By doing the same list for the top 20 users (Table 3) for betweenness it is possible to state that teachers are the users with higher level for this metric. Eight of the first ten users are teachers and the

first two are, also, members of the R&D team of SC, achieving levels considerable higher than the rest of all users. On the other hand, the eigenvector list shows that only one teacher (3rd) is part of the top 20 users. Closeness top list presents a more balanced distribution between teachers and students. With an exception of the top user (a teacher member of the SC R&D team) the remaining list only varies between 1.84 and 2.04. 13 users are students but, nevertheless, teachers tend to occupy the top places.

Table 3 Top 20 users for degree, closeness, betweenness and eigenvector

	Degree		In-Degree		Out-degree		Closeness Centrality		Betweenness Centrality		Eigenvector Centrality	
1	T01	986	T01	427	T01	559	T01	1.44	T01	319972	S05	1,0000
2	T02	584	T02	270	T02	314	T02	1.85	T02	126831	S16	0,9983
3	T03	520	S08	270	T04	294	S04	1.88	T07	79462	S14	0,9867
4	T04	505	S16	264	S04	274	T03	1,92	T05	44283	T03	0,9853
5	S01	466	S05	261	T03	269	T04	1,93	T08	40727	S01	0,9818
6	S02	448	S01	261	T06	261	S32	1,94	S43	38955	S03	0,9785
7	S03	446	S14	257	S24	256	S02	1,95	T04	32895	S08	0,9684
8	S04	426	S11	255	S25	240	S24	1,97	T03	31566	S09	0,9680
9	S05	425	S09	254	S26	239	T06	1,98	S04	29390	S11	0,9658
10	S06	418	S17	254	S27	238	S01	1,99	T09	19453	S22	0,9406
11	S07	413	S03	252	S28	238	S35	1,99	S44	13424	S49	0,9353
12	S08	407	T03	251	S29	238	S36	1,99	S08	13105	S19	0,9330
13	S09	402	S18	251	S13	238	T05	1,99	T10	12267	S50	0,9317
14	S10	401	S19	244	S15	237	S37	2,00	S32	11259	S02	0,9285
15	T05	400	S12	243	S10	237	S38	2,01	S45	10481	S51	0,9234
16	S11	396	S20	241	S30	235	S39	2,01	S46	8548	S21	0,9224
17	S12	387	S21	235	S31	234	S40	2,01	S03	8317	S52	0,9208
18	S13	387	S02	234	S32	234	S41	2,03	S01	7693	S17	0,9195
19	S14	386	S22	233	S33	234	S12	2,04	S47	7550	S53	0,9193
20	S15	386	S23	232	S34	234	S42	2,04	S48	7523	S54	0,9163

4.4 Group structure

Group detection was made using the modularity algorithm. In Fig. 5, the nodes' color represents the community each one belongs to and its size represents the weighted degree. Table 4 shows the detailed constitution of each community.

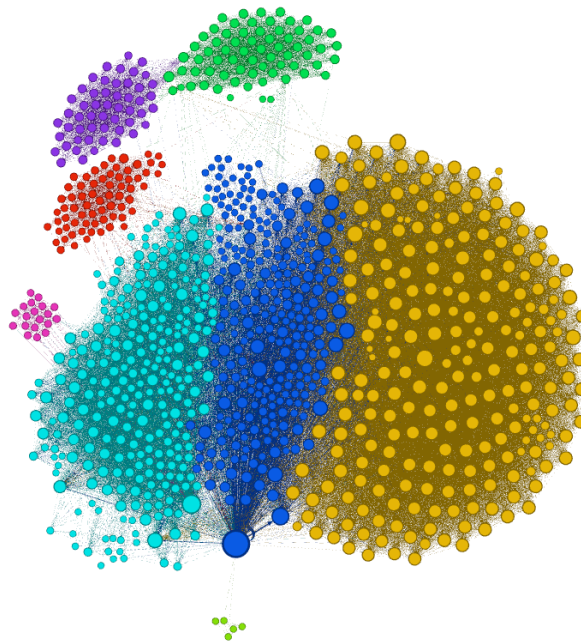


Figure 5 Sociogram with detected communities

Table 4 Communities constitution

	Color	Nodes	%
Community 0	Blue	276	28%
Community 1	Magenta	19	2%
Community 2	Cyan	275	27%
Community 3	Yellow	239	24%
Community 4	Green	5	0%
Community 5	Red	67	7%
Community 6	Purple	52	5%
Community 7	Light Green	68	7%

5 CONCLUSIONS AND FUTURE WORK

5.1 Conclusions and limitations

The results obtained through the SNA analysis made to the DeCA network clearly echo the scientific areas organization in the Department and especially their affinity and usage levels of the SAPO Campus platform.

Typically, this platform is more used by students in the Communication area, especially by undergraduate and masters' students. The usage pattern of the platform is visible in the actual results because they reveal that the majority of the network's actors have weak relations with other actors and that, in general terms, they have no distinct influence on others.

These results are aligned with the perception of use of the SC platform that we have as professors in the Department. In general terms, the students tend to use the platform to academic-related work and do not value so much its social and relational features.

Another important conclusion related with the aforementioned results is that SNA uncovers the importance, in this particular network, of some teachers in the promotion and dynamism of the network. The values obtained in the eigenvector measures and also in the degree and betweenness measures clearly reveal that there are some actors that achieve very interesting values in these analysis dimensions and that those actors are teachers.

As a method to analyze the actors, relationships and overall dynamism of a given network, from our point of view SNA is an approach that clearly benefits from a complementary admission of qualitative data in the overall analysis.

Although the overall picture that results from the SNA gives an important snapshot of the network, the actual comprehension of its nuances and causes it is enriched through data triangulation with other sources and types of data.

Moreover, the interpretation of the SNA results is also dependent of the actual knowledge that we have from the community dynamics, actors and relationships.

5.2 Future work: SAPO Campus SNA tools

The research work presented in this paper is framed by a work package of the GAMILearning project. The main aim of this particular work package is to develop a SNA export tool for the SAPO Campus platform that can provide, in an easy and quick way, relevant data regarding the aforementioned analysis dimensions to researchers and practitioners.

We believe that the simplification of this process of data gathering and preparation can promote the use of these metrics in terms of network analysis, assessment and continuous improvement.

To the specification of this data export tool it is paramount that all issues, ranging from the actual SNA theory to the technological issues of the SC platform, are correctly assessed and analysed in order to produce a tool that gives users with relevant and reliable data.

This work is a first research effort, clearly inspired by the SNA work initiated in Pais [14] PhD project.

The application of these metrics in DeCA has allowed us, as researchers, to confront our tacit and qualitative knowledge of the network (built in previous studies that resorted to different processes of data gathering and analysis with other techniques and instruments) with the quantified data that results from SNA analysis.

This confrontational strategy has allowed us assess the reliability of data gathered through SNA in a known context and extrapolate its applicability and utility for other SC contexts of use.

Thus, although the results were generally promising, they clearly indicate that the export process is still very intricate and that the SNA export tool to develop in SC should allow further data parametrization by the user, in order to select the relevant data for a given analysis.

ACKNOWLEDGEMENT

This work is part of the GAMILearning - Games for Media and Information Literacy - project (ref: UTAP-ICDT/IVC-ESCT/0020/2014) project funded by FEDER funds through the Operational Programme for Competitiveness Factors – COMPETE and National Funds through FCT – Foundation for Science and Technology (Portugal).

REFERENCES

- [1] Almeida, S., Pedro, L., Santos, C. (2014). Learning to be creative and being creative to learn with SAPO Campus: The development of a theoretical framework. *International Journal of Technology Enhanced Learning*, Vol. 6, N. ° 1, 65-77. DOI: 10.1504/IJTEL.2014.060023
- [2] Pedro, L., Santos, C. Almeida, S., Ramos, F., Moreira, A., Almeida, M., Antunes, M.J. (2014). The SAPO Campus recommender system: a study about students' and teachers' opinions. *Research in Learning Technology* 2014, 22: 22921 - <http://dx.doi.org/10.3402/rlt.v22.22921>.
- [3] Pedro, L., Santos, C., Aresta, M., & Almeida, S. (2015). Peer-supported badge attribution in a collaborative learning platform: The SAPO Campus case. *Computers in Human Behavior*, 51, 562–567. <http://doi.org/10.1016/j.chb.2015.03.024>
- [4] Santos, C., Almeida, S., Pedro, L., Aresta, M., & Koch-Grunberg, T. (2013). Students' Perspectives on Badges in Educational Social Media Platforms: The Case of SAPO Campus Tutorial Badges (pp. 351–353). Presented at the 2013 IEEE 13th International Conference on Advanced Learning Technologies (ICALT), Beijing, China: IEEE. <http://doi.org/10.1109/ICALT.2013.108>

- [5] Pais, F., Santos, C., Pedro, L. (2014). Innovation, Knowledge and Sustainability with PLEs: an Empirical Analysis from SAPO Campus Schools Pilots. *Journal of Literacy and Technology*, volume 15 number 2, 218-249.
- [6] Hanneman, R. and Riddle, M. (2005). *Introduction to Social network methods*. Riverside, CA: University of California, Riverside.
- [7] Izquierdo, L. & Hanneman, R., 2006. *Introduction to the formal analysis of social networks using mathematic*. Riverside, CA: University of California, Riverside
- [8] Fararo, T. J., 2000, April 21. Theoretical Sociology in the 20th Century. 2. Retrieved April 6, 2012, from <http://www.cmu.edu/joss/content/articles/volume2/Fararo.html>
- [9] Wasserman, S., Faust, K. (1994), *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- [10] Luque, J., Tyson, D., Lee, J., Gwede, C., Vadaparampil, S., Thomas, S., Meade, C. (2010). Using social network analysis to evaluate community capacity building of a regional community cancer network. *Journal of Community Psychology*, Vol.38, 5, 656-668.
- [11] Lusher, D., Robins, G., Kremer, P. (2010). The application of Social Network Analysis to team Sports. *Measurement in Physical Education and Exercise Science*, 14:211-224. Routledge: Taylor and Francis group.
- [12] Costa, C., Damásio, M. (2012). Elementary school goes online – making sense of peer-to-peer and peer-to-content interactions with SNA, MERJ (*The Media Education Research Journal*) 03:01.
- [13] Ehrlich, K., Carboni, I. (2005). *Inside Social Network Analysis*. IBM Technical Report 05-10.
- [14] Pais, F. (in press). *SAPO Campus Escolas: aprendizagem, ensino e pessoas em rede*. Unpublisehd PhD thesis (waiting for final examination). Universidade de Aveiro, Portugal.